DOCUMENT RESUME

ED 344 878 SP 033 789

AUTHOR Stacey, Dennis C.; And Others
TITLE Evaluation of the Effectiveness of the North Carolina Teacher Performance Appraisal System (TPAS).
PUB DATE Apr 88
NOTE 29p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 5-9, 1988).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Elementary School Teachers; Elementary Secondary Education; *Evaluation Criteria; *Evaluators; *Formative Evaluation; Job Performance; Secondary School Teachers; State Departments of Education; State Surveys; *Summative Evaluation; *Teacher Attitudes; *Teacher Evaluation
IDENTIFIERS *North Carolina Teaching Performance Appraisal

ABSTRACT

The North Carolina Teacher Performance Appraisal System (TPAS) embodies both formative and summative evaluation of teachers, and is based on multiple observations and collection of data on teacher performance. This study was conducted to report on a survey of the effectiveness of the TPAS as perceived by 639 evaluators (Es) and 2,732 teacher-evaluatees (Ts). Results are largely in the form of descriptive statistics which compare responses from evaluators and evaluatees, and TPAS-only and concurrent career development units. Overall findings indicate that respondents: (1) tended to evaluate the TPAS positively, particularly the summative evaluation component; (2) generally accepted criteria in the appraisal system; and (3) were positive about the processes of teacher observation and conferencing, but somewhat less positive about postconference followup. Es perceived postobservation conferencing and postconference follow-up slightly more positively than did Ts; and Ts tended to evaluate the formative aspects and the implementation of the TPAS less positively than did Es. However, all perceived the system as contributing at least moderately to improved teaching. (Author/LL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED344878

EVALUATION OF THE EFFECTIVENESS OF THE NORTH CAROLINA TEACHER PERFORMANCE APPRAISAL SYSTEM (TPAS)

Dennis C. Stacey
Barbara Kuligowski
David Holdzkom
North Carolina Department of Public Instruction
Raleigh, N. C. 27603-1712

A paper prepared for
the Annual Meeting of the American Educational Research Association
April 5-9, 1988; New Orleans, LA

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. Holdzkom

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

SO033789

ABSTRACT
Evaluation of the Effectiveness of the North Carolina Teacher Performance Appraisal System

Significant reforms efforts in North Carolina public education have occurred since the late 1970's, and in teacher evaluation, specifically, since 1980. A research-based teacher performance appraisal instrument was developed in 1983 (TPAI), and then supplemented by comparably standardized appraisal procedures. Together, these two components comprise the Teacher Performance Appraisal System (TPAS). This system embodies both formative and summative evaluation of teachers, and is based on multiple observations and collection of data on teacher performance. All participants in the TPAS have received 30 hours of training in Effective Teaching, and evaluators have received at least an additional 24 hours of training in using the TPAI.

The TPAS was pilot-tested in 40 North Carolina school units, 17 of which concurrently piloted a four-year career development program, during the 1985-86 and 1986-87 school years. This paper reports on a survey study of the effectiveness of the TPAS as perceived by 639 evaluators and 2732 teacher-evaluatees in 35 of the 40 pilot units, representing almost 100 percent participation from these units. Respondents tended to have major teaching experience at the elementary and high school levels, and more than ten years experience in their current jobs. Matched surveys for evaluators and evaluatees measured their perceptions of the structure, process and outcomes of the six major steps comprising the TPAS.

Results for this report are largely in the form of descriptive statistics which compare responses from evaluators and evaluatees, and TPAS-only and concurrent career development (CDP) units. Overall, respondents tended to positively evaluate the TPAS, particularly the summative evaluation component. Very little revision of the eight job functions and 38 practices, or the six-point rating scale which comprise the TPAI was indicated, suggesting general acceptance of the criteria of the appraisal system. Respondents were also generally positive about the processes of teacher observation and conferencing, and somewhat less positive about post-conference followup. All perceived the system as contributing at least moderately to improved teaching. Evaluators (Es) perceived post-observation conferencing and post-conference followup slightly more positively than teachers (Ts), as well as ten proposed outcomes or benefits of the TPAS.

In sum, Ts didn't tend to evaluate the formative aspects or the implementation of the TPAS as positively as Es. This less positive evaluation may be attributable to the formative aspects of the system being less developed than the summative. These differences were magnified in CDP units, meaning that CDP Es tended to more positively evaluate the TPAS than Es of TPAS-only units, while Ts of TPAS-only units tended to do so more than Ts of CDP units. Hypotheses for this apparent interaction between one's role and pilot unit include possibly greater technical sophistication of CDP participants, compounded by greater stakes and expectations or different motivations of CDP participants. Remaining issues focus on how to close the gap between Es' and Ts' experiences of the TPAS, particularly with regard to the formative aspects of the system, and where designated career development opportunities are available.

EVALUATION OF THE EFFECTIVENESS OF THE NORTH CAROLINA TEACHER PERFORMANCE APPRAISAL SYSTEM (TPAS)

D. C. Stacey, B. Kuligowski & D. Holdzkom
NC Department of Public Instruction; Raleigh, NC 27603-1712
April, 1988

The practice of evaluating teachers, while not new, has received renewed emphasis in the wake of the national realization that reforming education must become a high priority if America is to maintain its preeminent position in the global community. While it is commonplace to credit A Nation At Risk (National Commission on Excellence in Education, 1983) with succinctly stating the case for reform, it is probably more accurate to acknowledge that a number of individuals and groups were busy laying the groundwork for the reforms well before publication of the Commission's report. Since 1978, for example, the General Assembly of the State of North Carolina has been systematically addressing the needs--both in policy and fiscal arenas--for education improvements with respect to facilities, texts and text selection procedures, certification of educators, salary and benefits improvements, and educator evaluation.

An early outcome of this attention was the development of a consensus-based instrument for teacher evaluation. While recognizing the need for teacher evaluation, educators in the state also recognized in 1978 what Lortie had noted in 1975: "Research in education--until very recently a small-scale affair--has concentrated on learning rather than on teaching..." (Smith and Geoffrey, 1968, in Lortie p. 70). Thus, research-based knowledge of teaching skill was not available to inform the creation of evaluation systems. Instead, an elaborate process of collecting consensual views on teachers' responsibilities was undertaken, with invited participation from teachers, principals, and superintendents. These practices were then subjected to a factor analysis and an instrument for teacher evaluation resulted. The instrument was implemented in 24 school units and reliability studies were conducted (Inman, 1982). Not surprisingly, Inman found that principals tended to rate all teachers similarly on a number of dimensions. Moreover, there were serious disagreements about quality of performance when multiple evaluators observed a given teacher.

Further, a review of the evaluation procedures revealed that summative evaluation was the single goal of the entire system. While summative evaluation provides information for a school system to grant tenure, award merit pay, or make assignments to schools and classes, it does not provide information to stimulate or support changes in teaching performance. Unfortunately, even for summative purposes, the evaluation system embodied serious problems: (1) the criteria or standards themselves were not all clearly understood by all observers; (2) procedures for data collection and analysis varied from school to school; and (3) use of multiple observers was not required.

A second effort was deemed necessary. First, the standards for evaluation were re-examined. Through a contract with a group of university-based researchers, a thorough review of the research literature was undertaken (Group for the Study of Effective Teaching, 1983). Criteria for the literature review specified that a study had to meet a number of conditions to be included:

1) The study had to focus on one or more teaching practices that were observable.

2) The teaching practice had to be alterable.

3) The consequences of the practice had to be related to increased student achievement or time-on-task.

4) The research study had to be empirical.

The results of the search were reviewed by a panel of experts in effective teaching research and by a panel of teachers and administrators. This process isolated 28 specific teaching skills, that were grouped into five major teaching functions:

1. Management of Instructional Time
2. Management of Student Behavior
3. Instructional Presentation
4. Instructional Monitoring
5. Instructional Feedback

These five functions, then, comprise the core of the evaluation instrument. Moreover, three functions were added to reflect organizational values and norms, even though they did not meet the research criteria:

6. Facilitating Instruction
7. Interacting Within the Educational Environment
8. Noninstructional Duties

The resultant Teacher Performance Appraisal Instrument (TPAI) was comprised by these eight job functions, further defined into 38 practices.

Having established performance criteria, it became important to devise appraisal procedures that could be clearly understood and consistently applied. Appropriate procedures would also permit both goals of performance evaluation to be met: Evaluation should lead to ongoing opportunities for performance feedback and professional development, as well as summative judgments based on data collected during the year. Early on, the decision was made not to rate the teacher's performance after each observation. Rather, evaluators would be required to observe and code teachers' behaviors during classroom visits. As Peterson has pointed out,"...coding behaviors each time they occur...gives a good profile of the lesson observed and avoids the disadvantage, tedium, and large numbers of observations produced by category systems based upon three- or five-second intervals." (Peterson, 1985-86). It was, therefore, determined that while a numerical rating of each observation might be convenient for summative purposes, it would not facilitate formative evaluation.

The evaluation system, then, came to be characterized by: (1) multiple classroom observations, (2) observations that focused on a variety of teaching behaviors of high and low inference, and (3) documentation that would result in a set of narrative formative evaluation reports and a year-end quantified, summative evaluation report (Holdzkom, 1987). Obviously, the implementation of the system would depend upon competent, well-trained observers (Wise et al, 1984; Proctor and Lamkin, 1987; and Medley and Coker, 1987). In order to establish a basic understanding of the performance appraisal criteria, a 30-hour in-service training course on Effective Teaching was delivered to all persons who would

participate in the system, either as evaluators or evaluatees. In addition, a 24-hour course in the techniques of observation and evaluation was provided to evaluators, along with a six-hour course in development of professional growth plans. Finally, in February 1987, a ten-hour booster training program was delivered to evaluators.

Beginning, then, in 1985-86, North Carolina educators in 40 school districts were ready to field test the new performance appraisal system. The present study reports the results of an evaluation of this initial implementation of the North Carolina Teacher Performance Appraisal System (TPAS) in 35 of those 40 school districts from December, 1985, to June, 1987.

The survey study had three main objectives: (1) to constitute a large scale evaluation of the two-year TPAS pilot program, (2) to identify how the TPAS is affected by a related four-year pilot program in career development (CDP), and (3) to promote indirectly the formative evaluation and development aspects of the TPAS.

The TPAS has been described as having both formative and summative components. The TPAS encompasses and integrates these components by way of a prescribed evaluation cycle. The prescribed cycle occurs over the course of an academic year, and consists of at least three formative observations (perferrably one per each of the first three quarters of the year) that culminate in a year-end summative evaluation in which the teacher is rated, using a six-point scale, on the eight functions of the TPAI. The formative evaluations are based on formal (i.e. prescribed) observations and analyses of classroom teaching. Ideally, ongoing formative evaluation of a teacher can consist of more than three formal observations, and can include consideration of informal data.

The basic steps of the TPAS are: (1) Observation and documentation of teacher performance; (2) Coding of teacher performance according to functions and practices on the TPAI; (3) Analysis of documented and coded teacher performance; (4) Post-observation conferencing with the teacher to process the

analysis and structure performance improvement or professional development; (5) Post-conference followup to monitor and assist with implementation of conference agreements and recommendations; and, after the prescribed repetitions of the first five steps, (6) Summative evaluation and rating of the teacher on the TPAI. Steps three through five embody the more formative aspects of the system.

This appraisal system was designated to be pilot-tested for a two-year period by 23 North Carolina school units (hereafter referred to as the TPAS pilot units). In addition, a career development plan was concurrently piloted in 17 other North Carolina school units (hereafter referred to as the CDP pilot units). In these units, teacher career status, salary, and opportunities for enhanced professional responsibilities are based on attaining specified ratings on the TPAI. Considerably more technical assistance and training was provided to the CDP compared to the TPAS units relative to steps 1, 2, 3 and 6 of the TPAS.

METHODOLOGY

This survey study of the TPAS and CDP pilot units was undertaken by the Division of Personnel Relations of the North Carolina Department of Public Instruction, the office which also spearheaded the development of the TPAS and related pilot programs. The survey was conducted in April and May, 1987.

Sample

The Division of Personnel Relations invited the participation of all 23 of the TPAS pilot units, and 15 of 17 of the CDP pilot units. Two of the CDP units were not invited to participate because their unique programs were sufficiently dissimilar from that of the other 15 as to preclude meaningful comparisons. The decision of a unit to participate was officially that of the unit.

Sample selection from each school unit followed a standard format. A contact person in the central office of each unit distributed surveys to: (1) all central office staff who, in the contact person's judgment, were meaningfully involved in the TPAS, and (2) every administrator-evaluator of every school in the unit; school principals, in turn, distributed surveys to: (3) approximately four randomly-selected teachers for each evaluating administrator, and (4) one media specialist and one student services specialist (e.g. counselor) for approximately every 30 staff members of a school. Evaluators in CDP units include a small cadre of observer-evaluators,teachers who are trained to assist in the implementation of the TPAS on a rotational two- or three-year basis. In short, the desired sample included all administrators of the unit involved with the TPAS--in effect the population of evaluating administrators--and the equivalent of approximately one-fifth to one-sixth of the average evaluating administrator's caseload (i.e. four teachers).

Units which elected to participate in the survey included 22 of the 23 TPAS pilot units, and 13 of the 15 CDP pilot units. Based on the total number of surveys that were distributed, an estimated return rate of attempted surveys was 95 percent. Based on the number of surveys distributed to units which ultimately participated in the study, the return rate approximated 100 percent. An insignificantly small number of surveys was returned as unusable--i.e. not attempted, drastically incomplete, or unable to be completed or adjusted for inclusion in the analysis. In all, usable surveys were returned by 639 evaluators and 2732 evaluatees--the vast majority of whom were classroom teachers.

Background information collected on respondents revealed that they tended to most report elementary and high school teaching experience, and more than ten years experience in their current job. Thirty-three percent of school administrators and 41 percent of observer-evaluators reported high school as

their major teaching experience, while approximately 38 percent of both reported elementary school experience. Of responding teachers, 50 percent indicated elementary school as their major teaching experience, followed by 26 percent who indicated high school. In terms of years of experience in one's present job, "more than ten years" was reported by 76 percent of observer-evaluators, 35 percent of school administrators, and 56 percent of teachers. The remaining school administrators were approximately equally distributed among experience categories of less than two years, three to five years, and six to ten years. In short, most responding evaluators were characterized by major teaching experience at the high school level, and more than ten years experience in their current job. Most responding teachers reported major teaching experience at the elementary followed by high school level, as well as more than ten years experience in their current job. Finally, school administrators indicated that they devoted almost half of their jobs to TPAS responsibilities.

The Survey

In an attempt to thoroughly and fairly capture the outcomes of a complex and critical pilot experience of two years duration, a relatively comprehensive and elaborate survey was constructed. To determine the effectiveness of the TPAS from both evaluator and evaluatee perspectives, companion or matched versions of the survey were designed. Both the evaluator and evaluatee versions of the survey contained two to three items which served to identify respondents as to role and experience, and 62 major items that comprised the dependent variables of the study. In addition, the evaluator version included 11 major items which delved into Es' experience with the TPAS; six major items that tapped outcomes of any co-evaluation experiences they may have had; and eight other major items, mostly relating to the involvement of non-principal evaluators in the TPAS. Response formats for most survey items involved (1) three-

to four- alternative, continuous variable choices (e.g. "little/none" or "never" to "very much" or "often"), (2) percentage attributions that were to add to 100 percent, and (3) frequency and time estimates (evaluators' version only). It is estimated that the evaluator survey took roughly one hour to complete, and that the evaluatee or teacher survey took roughly one half-hour to complete.

Both versions of the survey sought to measure respondents' perceptions of the structure, process and outcomes of the six steps of the TPAS described earlier. Survey items on the structure of summative evaluation required that respondents evaluate the content of the TPAI (i.e. the eight job functions and the six-point rating scale). Respondents were also asked to assess up to ten possible outcomes or benefits of the TPAS which impact upon teachers', evaluators' and principals' jobs.

RESULTS

Building on the previously stated purposes and methodology of this study, the following equivalent of an experimental design for the study will serve as a framework for the presentation of findings. Since the survey was conceived to determine the effectiveness of the teacher performance appraisal system (TPAS), the measured effectiveness of the TPAS can be construed as the dependent variable. This effectiveness, in turn, was primarily studied as a function of: (1) one's role in the TPAS (evaluator or, teacher), (2) the type of pilot unit in which the TPAS was used (TPAS or CDP unit), and (3) the unique characteristics of the various CDP units. These potential determinants of the effectiveness of the TPAS can thus be construed as independent variables. For purposes of this paper, only the effects of first two of the proposed independent variables will be emphasized.

Moreover, the effects of one's role in the evaluation system and the type of pilot unit in which one participated have been largely analyzed for this paper in the form of descriptive statistics (i.e. means and frequency distributions of select dependent variables). This level of analysis was deemed appropriate for this study, which is conceived as an exploratory study of experimental personnel programs. Such analyses will be reported for the major independent groups or variables already mentioned, as well as for a further differentiation of the "role" variable according to subgroup or moderator variables: (1) evaluator effects have been further studied according to evaluator type (school adminstrator or observer-evaluator (O/E)), (2) evaluatee or teacher effects have been further studied according to the grade level of one's major teaching experience, and (3) both evaluator and evaluatee effects have been further studied as to extent of one's experience in his/her present job.

In a brief departure from this descriptive statistics approach for analyzing the results of this study, nine t-tests of independent group means were computed for select dependent variables in an attempt to sample the statistical significance of some of the results. Each t-test compares the role variable of evaluator and teacher on: (1) four dependent variables related to post-observation conferences, (2) two dependent variables related to post-conference followup, and (3) three dependent variables related to possible outcomes of the TPAS. Since the results of these preliminary tests were uniformly highly significant, a subsequent multivariate analysis of the relationship of all of the independent variables to major dependent variables is planned as the next phase of this investigation.

Overall Effects of the TPAS

Before reporting results for each of the independent variables, it should be stated that the overall evaluation of the TPAS--regardless of one's role or affiliation--is largely in a positive direction. This is particularly true

when one considers the summative aspects of the TPAS. It is the summative process that has been emphasized in the early stages of implementation (e.g. training for using the teacher performance appraisal instrument on which to rate teachers).

When respondent groups evaluated the formative aspects of the system, opinions about its effectiveness were more divergent. This is not surprising since less attention has been paid to formative evaluation or analyses, and even less to formative development of teachers.

Observation

The quality of teaching associated with formal observations is perceived as average or better. Both evaluators (Es) and teachers (Ts) perceive at least some relationship of later observations of a given teacher to observations conducted earlier in the year. Where multiple observers have evaluated a teacher's performance, both Es and Ts perceive at least "some" agreement among these observers.

Post-Observation Conferencing

Evaluators and Ts perceive that the amount of time each has for conferencing is between "just enough" and "more than enough." They also agree that between "just enough" and "more than enough" concrete evidence is used to draw conclusions about a teacher's performance. Also, approximately three-fourths of both Es and Ts report being able to come to terms with each other's perception of the teacher's performance "to a large extent."

Post-Conference Followup

The findings that relate to what happens after or between observations of a teacher are consistently least positive of all. First, in response to the closing item of the survey which requested an overall effectiveness rating of the five major steps of the TPAS, both Es and Ts clearly rated the effectiveness of "post-conference monitoring and followup" the least effective of all steps. Second, when asked to evaluate the contributions of O/Es and designated evaluators to each step of the TPAS, Es designated "post-conference monitoring and followup" as the area where these evaluators are least effective. Finally, when Es and Ts were asked to evaluate the adequacy of release time and on- and away from-the job opportunities for professional development of Ts, both rated these opportunities, on the average, as less than "just about right."

At the same time, both Es and Ts perceive that Ts request and subsequently receive Es' assistance, and implement conference recommendations to at least "some" degree. In addition, Es and Ts see as "just about the right the amount" of time available for Ts to pursue improved performance between evaluations.

Summative Evaluation

Respondents are generally very accepting of the functions of the Teacher Performance Appraisal Instrument (TPAI). When asked about the extent to which job functions should be revised, Es' and Ts' mean responses fall between "not at all" and "only slightly." This relative lack of desired revision of the TPAI functions suggests a unified acceptance of the criteria upon which summative evaluations are made.

While general acceptance of the rating scale of the TPAI by Es and Ts is also reflected in the findings, the parties differ in some minor respects. The current six-point rating scale is clearly the size of scale that Ts prefer. Es favor a five-point scale slightly more than the six-point scale. Revisions of

the definitions of rating scale positions are desired on the order of "none" to only "slight," with the latter applying most to ratings 5 (well above average) and 6 (superior). So, again, there is notable acceptance by Es and Ts of a six-point rating scale, although some clarification of the criteria for ratings 5 and 6 is desired.

Finally, over 80 percent of both Es and Ts view both the information upon which principals base their summative evaluation and Ts' acceptance of principals' summative evaluation as "largely adequate."

TPAS As a Whole

Both parties view the TPAS as contributing at least "moderately" to improved teaching, and as least "moderately effective" in terms of each of the five major steps (excluding "Coding of observation data") which comprise the system, including the more formative post-conference followup.

Role of TPAS Participants: Between-and Within-Group Comparisons for Evaluators and Teachers

Observation

While both parties believe that the quality of teaching is better during announced than unannounced observations (and Es believe this more than Ts), they disagree on the quality exhibited during unannounced observations, with Ts perceiving better teaching during unannounced observations than Es. Moreover, with regard to subsequent observations being related to earlier identified teacher needs, observer-evaluators (O/Es) tend to report more of such continuity than school administrators (SAs). Finally, this continuity of observations varied directly with administrators' experience in the role of teacher-evaluator, with administrators with 10 or more years of experience reporting the most continuity.

Post-Observation Conferencing

First, Es more than Ts believe that Ts have at least adequate time to present their side during conferences ($t=6.343$, $p<.001$), with O/Es tending to

exhibit slightly more of this belief than SAs. Es also believe more strongly than Ts that conclusions about a T's performance are based on concrete evidence ($t=11.823$, $p<.001$). Again, O/Es express slightly more of this belief than SAs, and elementary Ts express more than Ts of other levels. Where differences in perception of a teacher's performance occur, Es more than Ts see such differences as being largely resolved. The most experienced Ts (more than ten years) perceived the least resolution. On the matter of consistency of evaluation, Es more than Ts perceive (1) more relationship between conferences in the evaluation cycle ($t=5.031$, $p<.001$), and (2) more agreement about the teacher's performance where more than one observer is involved ($t=7.934$, $p<.001$).

Post-Conference Followup

Compared to Ts, Es see themselves as much more likely to initiate followup with Ts to both (1) monitor Ts' implementation of conference recommendations ($t=11.143$, $p<.001$), and (2) assist Ts in this process. With respect to monitoring, tendencies were noted for the most experienced Es to report doing the least monitoring, and for elementary and middle-school Ts (especially elementary) to perceive a greater prevalence of monitoring than secondary school Ts (especially junior high school Ts). When it comes to Ts seeking help to implement conference recommendations, elementary and middle school Ts (especially elementary) tend to report seeking such help more than secondary school Ts (especially junior high school Ts). In a similar fashion, elementary school Ts tend to perceive Es as being more helpful to them in implementing conference recommendations than other Ts. Ultimately, Ts saw themselves as carrying out conference agreements or recommendations to a greater extent than did Es ($t=3.733$, $p<.001$).

Summative Evaluation

Both Es and Ts desire a reduction in the number of required formal observations of Ts, and a corresponding greater emphasis on self-evaluation, followed

by peer- and consumer-evaluation. Whereas Es almost equally prefer a five-point and the current six-point rating scale, Ts prefer the six-point scale by almost two to one. With respect to pre-summative feedback on the teacher's performance, Es more than Ts regard the feedback offered by the current formative evaluations as sufficient. On the other hand, Ts prefer some kind of ratings before year-end summative evaluation. While more than three-quarters of Ts see summative evaluations as "to a large extent" based on adequate information, a slight tendency is noted for middle-school Ts to most endorse this adequacy and for high school Ts to least endorse it. Es tend more than Ts to perceive principals' summative evaluations to be accurate indicators of a teacher's performance.

Overall TPAS Benefits

Of ten proposed benefits of the TPAS, Es endorse all ten more than do Ts. Whereas Es attribute between "moderate" and "significant" benefit or effectiveness to eight of ten possible outcomes of the TPAS, Ts attribute as much to only one--improved teaching. Certainly this is one of the most important intended outcomes of the whole pilot effort. The other benefits of the TPAS which Es favored more than Ts were better staff development, clearer guidelines for improving teacher performance ($t=6.507$, $p<.001$), better teacher conferences ($t=6.411$, $p<.001$), more objective performance appraisal ($t=11.214$, $p<.001$), and greater ease in making employment decisions. Ts tended to attribute between "slight" and "moderate" benefit or effectiveness to these outcomes.

A further look at "better teaching" as an outcome of the TPAS reveals that the least experienced Es (less than two years) tend to be less endorsing of this outcome than Es of other experience levels, while Es with from six to ten years of experience tend to be the most endorsing. As far as teacher differences, middle-school Ts tend to be slightly more endorsing of the "better teaching" outcome than Ts of other grade levels.

Both Ts and Es endorse "job satisfaction" for Ts the least of all of the ten possible TPAS outcomes, and for Ts this endorsement approaches only a "slight" benefit. Elementary and middle school (especially middle-school) Ts tend more than secondary school Ts (especially junior high school Ts) to endorse job satisfaction as a benefit or outcome of TPAS. The most experienced Ts (more than ten years) tend to least endorse job satisfaction as a possible benefit of the TPAS.

With respect to the TPAS offering clear guidelines for improved teacher performance, disproportionately more O/Es than SAs see this outcome as tending toward "significant." Among Ts, there is a tendency for the most experienced (more than ten years) to be least likely to endorse "clearer performance guidelines" as a benefit.

Finally, among Es, the most experienced (six to ten years and more than ten years) tend to see "fairer and more objective and accurate evaluations of Ts" as more of a significant outcome of the TPAS than do less experienced Es.

In summary, some notable differences in Es and Ts appraisals of the TPAS were identified, many of which have to do with the critical formative aspects of the system. While both Ts and Es report that the TPAS generally operates as intended to at least "some" or a "moderate" degree, Ts do not tend to be as satisfied as Es with aspects of post-observation conferencing, post-conference followup, and ultimately summative evaluation. Consequently, Ts do not endorse possible benefits of the TPAS to the same degree as do Es.

In addition, some interesting within-group trends were noted for Es and Ts. In some respects, O/Es more than principals perceived themselves as fostering the following formative aspects of the TPAS: Continuity of conferences, adequate time for Ts to represent themselves in conferences, analyses of teacher performance based on concrete evidence, and clear guidelines for improving teacher performance.

Elementary and middle school Ts and, to a lesser degree, less experienced Ts tend most to perceive and foster the following formative aspects of the TPAS: Analyses of performance based on clear evidence, resolution of differences in perception of a teacher's performance, post-conference monitoring and coaching from Es, and clear guidelines for teacher improvement.

Type of Pilot Unit: TPAS and CDP Units

Due to the greater investment of planning and resources in the appraisal system for the CDP compared to the TPAS pilot units, the two types of units will first be compared and profiled as to evaluator differences. In a word, CDP Es have approximately 50 percent more experience than TPAS Es with every phase of the six-step TPAS outlined earlier. To further support the earlier contention that the formative evaluation aspects of the TPAS need to be reinforced, both groups of Es report having approximately half of the overall experience with post-conference monitoring as they do with other aspects of the process. Moreover, differences between evaluators of the two types of units are noted with regard to their respective experiences with formative observation data analysis (FODA) and post-conference monitoring. While CDP Es reported devoting an average of 60 minutes to preparing a FODA, TPAS Es reported a comparable figure of 42 minutes. Moreover, whereas CDP Es required 19 weeks and 28 to 41 FODAs to feel comfortable with this step, TPAS Es required 14 weeks and 14 to 16 weeks, respectively. On the other hand, TPAS Es reported devoting an average of 52 minutes to each teacher for post-conference monitoring, compared to CDP Es' average estimate of 35 minutes. Overall, while both groups tend to find all steps of the TPAS "quite manageable," CDP Es report more overall manageability than TPAS Es.

Evaluator-Teacher Differences Within Pilot Units

Before examining these differences, it is important to note the overall trend of these findings: While CDP Es tend to more positively evaluate the outcomes of the TPAS than do TPAS Es, Ts' evaluations of the TPAS tend to be more positive within TPAS than CDP units.

Post-Observation Conference

CDP more than TPAS Es tend to see their conclusions about teacher performance as more data-based. For Ts, on the other hand, TPAS Ts more than CDP Ts perceived conclusions about their performance as being more data-based, and also perceived: (1) greater continuity among their conferences, (2) more opportunities to participate in their conferences, and (3) greater agreement with their Es about their performance, including resolution of differing viewpoints.

Post-Conference Followup

When it comes to opportunites and resources for helping Ts to improve, CDP Es more than TPAS Es tend to see such assets as available in their units. However, TPAS more than CDP Es (1) see themselves as monitoring and assisting with post-conference improvement of Ts, and (2) see Ts as asking for and receiving such help.

Among Ts, it is the TPAS more than CDP Ts who see growth possibilities related to opportunities and resources for improvement available in their units.

Summative Evaluation

CDP principals tend more than TPAS principals to see summative evaluations of Ts as both based on adequate information, and as being valid indicators of teacher performance.

TPAS As a Whole

When asked to assess the relative benefits of TPAS from a list of ten possible outcomes, CDP Es attach more significance to nine of ten options than

do TPAS Es. The only benefit that TPAS Es rate higher than CDP Es is "improved communication" between school administrators and Ts. Similarly, with respect to the six steps of the TPAS, CDP Es attribute more effectiveness to all of the steps, except for post-conference monitoring. This finding may parallel the earlier one that, on the average, TPAS Es report spending more time than CDP Es on post-conference followup of Ts.

In contrast to the above findings, TPAS Ts, as compared to CDP Ts, attribute more benefits to the TPAS, and more effectiveness to the six-step process in every instance where a difference was apparent.

DISCUSSION

The results reveal that, overall, both Es and Ts favorably regard the TPAS. Mean responses of both groups indicate a general adequacy, acceptance and support of the TPAS. The most notable differences between Es and Ts relate to post-observation conferences, post-conference followup, summative evaluation, and benefits accruing from the TPAS. While both groups are generally satisfied with the TPAS, Es' satisfaction exceeds that of Ts for several aspects of the TPAS. Es regard more positively than Ts the adequacy of data used in evaluations, as well as the accuracy of those evaluations in capturing teacher performance. They also believe more than Ts that more on-going, formative evaluation and development occurs, as well as greater resolution of differences in perceptions of teacher performance.

When evaluation of the TPAS is examined as a function of participation in a particular pilot program (CDP or TPAS) the same trends emerge. However, the differences between Es and Ts in their evaluations of the TPAS are greater in CDP units. Stated differently, while Es in CDP units assess the outcomes of the TPAS (especially outcomes related to summative aspects of the system) more

favorably than Es in TPAS units, more positive regard for the TPAS is expressed by TPAS Ts than CDP Ts. This greater disparity of Es' and Ts' opinions in CDP units is of particular interest because Es and Ts in CDP units have received more training and technical assistance than those in TPAS units, and they have additional performance appraisal resources in the form of observer-evaluators.

Explanation of the Results

Overall Responses to the TPAS

All groups of respondents desire little revision of the job functions and practices, and the rating scale that comprise the TPAI. In addition, both Es and Ts report general satisfaction with the established procedures of the appraisal system. These results are consistent with earlier studies conducted by the North Carolina Department of Public Instruction, CDP units, and a state teachers' organization. In short, there appears to be an evolving satisfaction with the TPAS as it is designed and spelled out on paper.

Differences between and among groups in relative satisfaction with the TPAS appear to stem from implementation of the system. Consistent with this proposed explanation, differences between both Es and Ts as well as CDP and TPAS units are a matter of degree, not substance. No unique effects are apparently associated with the CDP.

Implementation issues point to the need for mutual and consistent interpretation and application of performance standards by both Es and Ts. What appears neccessary to ameliorate the differences between Es and Ts noted here is not changes in the design and intent of the system, but (1) continued refinement of performance-related planning, judgments, and improvement on the part of both parties, as well as (2) greater and more effective communication and collaboration between the parties as they try to do more justice to the ongoing formative aspects of the system.

It is not surprising that Es and Ts evaluate the effectiveness of the most formative part of the TPAS (i.e. post-conference monitoring and followup) as the least effective of all, as well as differ in their perceptions of the formative aspects of the TPAS. Formative evaluation is a relatively new and undeveloped practice, not formally incorporated into teacher appraisal in North Carolina until 1985. The active participation of Ts intended by this process represents a major shift from the more passive role implied by or inherent in the more traditional summative model. In addition, evaluators are being called upon to become more active and effective in prescribing, monitoring, and supervising or coaching improvements in teacher performance. In all likelihood, North Carolina educators have responded to the complex demands of the TPAS by tending to sequence or prioritize their learning of and accommodation to the system, starting with the "more pressing", more familiar or more easily learned summative aspects of the system. Thus, educators are still coming to terms with the notion of shared and ongoing responsibilities for analyzing, discussing, and working to uniformly improve teacher performance in an atmosphere of mutual trust and respect.

While implementation issues are not to be dismissed lightly, they call for different responses and solutions than design problems. What seems to be in order is a fine-tuning not an overhaul of the system. The TPAS is an elaborate system that calls for well-developed skills in observation, analysis, improvement-planning, and written and oral communication. It is a system that requires participants to allocate time and energy--precious resources in today's world of education--if it is to succeed. Evaluators and Ts agree that improved teaching is one outcome that has already been realized from the effort. It seems that continued improvement of teaching alone would justify the effort required to continue to fine-tune the system.

CDP Results

As noted, Es and Ts in CDP units differ more than those in TPAS units in their evaluations of the TPAS. It may be that CDP enhances or compounds differences in Es' and Ts' perceptions of the TPAS. Two hypotheses are offered in this regard.

The first hypothesis is suggested by the additional training and technical assistance that CDP units have experienced. One outcome of this greater investment may be that CDP Es and Ts are more sophisticated and rigorous in their application of TPAS. They may, therefore, be able to make finer discriminations about T-performance and improvement. However, to the extent that Es and Ts have developed and applied these skills independently of each other, it is more likely that they will differentially evaluate the teacher's performance and be unable to resolve differing perceptions of performance. This situation may be compounded if Es and Ts are developing their respective sophistication at different rates.

While this hypothesis seems to explain the finding that CDP Es surpass TPAS Es in their beliefs that their conclusions about a teacher's performance are based on adequate data and accurately capture teacher performance, it is not supported by comparable results for Ts: TPAS more than CDP Ts believe that conclusions abut their performance are data-based and that the TPAS has produced better guidelines for improvement and more objective evaluations. Therefore, another factor besides increased sophistication in use of the system seems to be operating for Ts.

Accordingly, a second hypothesis is proposed. Because the CDP offers the potential for increased salary and expanded professional opportunities, experiences with the TPAS in CDP units are associated with heightened stakes and greater expectations. Consequently, it may be more important for Es and Ts to be right or believe they are right. With greater stakes, both parties have a

greater investment in their own perceptions and judgments about T performance, which, as suggested above, may already differ. If the emphasis is on who is right, reaching agreement becomes more difficult. Participants may resort to personally biased, data-justified, rather than mutually-derived, data-driven positions.

In non-career development units, on the other hand, when no additional benefits are connected to performance appraisal outcomes, it may be less risky and easier to offer and agree with formative conclusions and suggestions, no matter who offers them. It may also be easier to agree on summative ratings since nothing material is gained or lost through compromise. Thus, TPAS Ts may not be as motivated as CDP Ts to differ with or refute evaluators' conclusions about their performance since they don't qualify for the significant rewards and status opportunities which are available in CDP units. Moreover, TPAS Es may not be as driven as CDP Es to insure that Ts truly earn their rewards and that their (the Es') decisions are justified.

Thus, the effects of increased technical skill for making judgments about teacher performance proposed by the first hypothesis could be exacerbated by the motivational factor proposed by this second hypothesis. The reality is that probably both increased and/or differential sophistication and heightened ego-involvement have contributed to CDP Es and Ts differentially valuing the TPAS. Accordingly, true differences in perceptions of teacher performance are more firmly held to and enhanced when status and rewards are at stake.

Unresolved Issues

As stated earlier, this investigation was undertaken as an exploratory activity to yield primarily descriptive information about the TPAS. The survey yielded a rich data base. This report touches on the primary themes emerging from the data. It also raises questions for further study. Some of the issues that remain to be addressed are set forth here.

Do all evaluators and teachers have the aptitude or the skills needed to successfully implement the TPAS? If the necessary aptitude is in short supply or the skills remain insufficiently developed, the formative evalaution goals of the system will not be attained, and the system as designed will fail. Minus an effective formative component, the TPAS would join the world of performance appraisal already replete with approaches to summative evaluation. Clearly, then, the sophistication required to do justice to the system requires a continuation of the on-going training and technical assistance that have been offered. Is this approach to developing the necessary skills enough? If so, what types of training and technical assistance are most effective? If not, what can be done to develop and refine requisite skills?

The resources which are necessary to comprehensively and successfully operationalize this system are limited. Should they be selectively expended on the most ready and willing educators (e.g. the most proven or experienced evaluators, or the most accessible and receptive teachers, such as elementary or middle school teachers or relatively less experienced teachers)?

Do teachers who receive high ratings support TPAS more than teachers who receive low ratings? This study did not ask respondents to report either formative or summative evaluation results for teachers. However, the general agreement on the acceptability of the functions, practices, and rating scale of the TPAI and other aspects of the system indicates that the TPAS embodies standards and procedures against which a majority of teachers are willing to be judged. This is a significant finding. What remains is to improve implementation and application of this system toward greater mutual satisfaction with performance appraisal and professional growth of all parties. Thus, the important question is not whether the ratings teachers receive influence their approval of the system, but whether teachers think the system as prescribed is fair and appropriate. The data reported here say that they do.

BIBLIOGRAPHY

The Group for the Study of Effective Teaching (1983). Teaching Effectiveness Evaluation Project: Final Report. Chapel Hill, NC: The University of North Carolina.

Holdzkom, David (1987, April). Appraising Teacher Performance in North Carolina. Educational Leadership, pp. 40-44.

Inman, William C. (1982). Reliability of the Teacher and Principal Personnel Performance Appraisal Instruments. Raleigh, NC: North Carolina State Board of Education.

Inman, William C. (1982). Reactions of Teachers, Principals, and Superintendents to 1981-82 Performance Appraisals. Raleigh, NC: North Carolina State Board of Education.

Lortie, Dan C. (1975). Schoolteacher: A Sociological Study. Chicago: The University of Chicago Press.

Medley, Donald M. & Coker, Homer (1987, October). "How Valid are Principals' Judgments of Teacher Effectiveness?" Phi Delta Kappan. pp. 138-140.

National Commission on Excellence in Education (1983). A Nation At Risk: The Imperative for Educational Reform. Washington, DC: U. S. Department of Education.

Peterson, Donovan. Developing Teacher Evaluation Systems With Potential for Increasing Student Performance. Educational Research Quarterly. (Vol. 10, No. 2. pp. 42-43).

Proctor, Thomas J. & Lamkin, Bill D. (1987, Winter). "Fairness in Teacher Evaluation." Capstone Journal of Education. (Vol. 7, No. 2 pp. 19-22).

Wise, Arthur; Darling-Hammond, Linda; McLaughlin, Milbrey W. & Bernstein, Harriet (1984). Teacher Evaluation: A Study of Effective Practices. Santa Monica, CA: The Rand Corporation.

FACTS AND FINDINGS
Survey Study of the TPAS
1986-1987

* Thirty-nine North Carolina school units participated in the field-test of the new TPAS.

* Almost 3400 surveys were analyzed in this study, representing 639 administrators and 2732 teachers. This represents about 95 percent of all initially distributed survey forms, and almost 100 percent return by participating units.

* All four response groups--Career Development evaluators, Career Development teachers, TPAS Pilot evaluators and TPAS Pilot teachers--were largely positive about virtually all aspects of the Teacher Performance Appraisal System.

* Evaluators and teachers agree that concrete evidence from observations formed the basis of post-observation conferences. More importantly, about three-fourths of evaluators and teachers reported that they agreed with one another's perception of the teacher's performance "to a large extent".

* All respondents agreed that the eight functions of the TPAI required no revision or only slight revisions. Also, teachers preferred the six-point scale by a large majority, while evaluators were divided between a five and six-point scale, with a slight preference for the five-point.

* More than 80 percent of both teachers and evaluators viewed the information on which summative evaluation was based as "largely adequate".

* Evaluators attribute moderate to significant benefit or effectiveness to several possible outcomes of the TPAS: improved teaching, better staff development, clearer guidelines for improving teaching, better conferences with teachers, more objective performance appraisal and greater ease in making employment decisions.

* No major differences were revealed for CDP compared to TPAS respondents, although differences of degree were noted between evaluator and teacher opinions in the two kinds of units.

* While summative aspects of the system appear to be well-established, increased attention to formative evaluation and development seems to be required.

N.C. TPAS SURVEY STUDY

MAJOR FINDINGS: MEAN DIFFERENCES

	Evaluator	(Unit Differences)	Teacher	(Unit Differences)
Post-Observation Conference				
* Adequacy of teacher time (3 to 1 scale, 1 is high)	1.44 >		1.60	(TPAS>CDP)
* Data-based conclusions (same as above)	1.37 >	(CDP>TPAS)	1.71	(")
* Relationship of multiple conferences (1 to 3 scale)	2.42 >		2.19	(")
* Resolution of differences: To a large extent	84% >		68%	(")
* Inter-rater agreement (3 to 1 scale)	1.12 >		1.31	
Post-Conference Followup				
* Monitor teacher-improvement (1 to 4 scale)	3.21 >	(TPAS>CDP)	2.71	(TPAS> CDP)
* Help teacher-improvement (1 to 4 scale)	3.14 >	(")	2.64	(")
* Teacher implementation of recommendations (1 to 3)	2.44 <		2.53	(TPAS>CDP)
TPAS Outcomes/Benefits (0 to 3 scale)				
* Better teaching	2.40 >	(CDP>TPAS)	2.02	
* Teacher satisfaction	1.64 >	(CDP>TPAS)	1.46	(TPAS>CDP)
* Teacher commitment	1.87 >	(")	1.75	(")
* Better staff development	2.14 >	(")	1.81	
* Better improvement guidelines	2.33 >	(")	1.99	(")
* Easier performance appraisal (PA)	1.94 >	(")	1.77	
* More objective PA	2.30 >	(")	1.77	(")
* Easier employment decisions	2.17 >	(")	1.85	
* Increased communication	1.92 >	(TPAS>CDP)	1.69	(")
* Better conferences	2.22 >	(CDP>TPAS)	1.88	(")
Summative Evaluation Ratings				
* Accurate indicator of overall performance	84% >	(CDP>TPAS)	70%	